(54) Title: VIDEO ENCODING AND DECODING METHODS AND CORRESPONDING ENCODING AND DECODING DE-
VICES

(57) Abstract: The invention relates to an encoding method applied to an input video sequence corresponding to successive scenes
subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of said scenes, a coded bit-
stream the content of which is described in terms of separate channels and constituted of encoded video data in which each data item
is described by means of a bitstream syntax allowing to recognize and decode all the elements of said content, said syntax comprising
an additional syntactic information provided for describing independently the type of temporal prediction of the various channels.
According to the invention, said additional information is a syntactic element placed at the slice level or the macroblock level in the
coded bitstream, and its meaning is either specific for each present channel or shared by all existing channels.

"VIDEO ENCODING AND DECODING METHODS AND CORRESPONDING
ENCODING AND DECODING DEVICES"

FIELD OF THE INVENTION

The present invention generally relates to the field of video compression
and, for instance, to the video coding standards of the MPEG family (MPEG-1, MPEG-
2, MPEG-4) and to the video recommendations of the ITU-H.26X family (H.261, H.263
and extensions, H.264). More specifically, this invention relates to an encoding method
applied to an input video sequence corresponding to successive scenes subdivided into
successive video object planes (VOPs) and generating, for coding all the video objects
of said scenes, a coded bitstream the content of which is described in terms of separate
channels and constituted of encoded video data in which each data item is described by
means of a bitstream syntax allowing to recognize and decode all the elements of said
content, said syntax comprising an additional syntactic information provided for
describing independently the type of temporal prediction of the various channels, said
predictions being chosen within a list comprising the following situations :

- the temporal prediction is formed by directly applying the motion field
sent by the encoder on one or more reference pictures ;

- the temporal prediction is a copy of a reference image ;

- the temporal prediction is formed by the temporal interpolation of the
motion field ;

- the temporal prediction is formed by the temporal interpolation of the
current motion field and further refined by the motion field sent by the encoder .

The invention also relates to a corresponding encoding device, to a
transmittable video signal consisting of a coded bitstream generated by such an
encoding device, and to a method and a device for decoding a video signal consisting of
such a coded bitstream.

BACKGROUND OF THE INVENTION

In the first video coding standards and recommendations (up to MPEG-4
and H.264), the video was assumed to be rectangular and to be described in terms of a
luminance channel and two chrominance channels. With MPEG-4, an additional
channel carrying shape information has been introduced. Two modes are available to

compress those channels : the INTRA mode, according to which each channel is
encoded by exploiting the spatial redundancy of the pixels in a given channel of a single
image, and the INTER mode, that exploits the temporal redundancy between separate
images. The INTER mode relies on a motion-compensation technique, which describes
an image from one or several image(s) previously decoded by encoding the motion of
pixels from one image to the other. Usually, the image to be encoded is partitioned into
independent blocks or macroblocks, each of them being assigned a motion vector. A
prediction of the image is then constructed by displacing pixel blocks from the reference
image(s) according to the set of motion vectors (luminance and chrominance channels
share the same motion description). Finally, the difference (called the residual signal)
between the image to be encoded and its motion-compensated prediction is encoded in
the INTER mode to further refine the decoded image. However, the fact that all pixel
channels are described by the same motion information is a limitation damaging the
compression efficiency of the video coding system.

SUMMARY OF THE INVENTION

It is therefore the object of the invention to propose a video encoding
method in which said drawback is avoided by adapting the way the temporal prediction
is formed.

To this end, the invention relates to a method such as defined in the
introductory part of the description and which is moreover characterized in that said
additional syntactic information is a syntactic element placed in said generated coded
bitstream and its meaning is specific for each present channel, said element being placed
at the slice level or at the macroblock level according to the proposed embodiment.

The invention also relates to a corresponding encoding device, to a
transmittable video signal consisting of a coded bitstream generated by such an
encoding device, and to a method and a device for decoding a video signal consisting of
such a coded bitstream.

DETAILED DESCRIPTION OF THE INVENTION

According to the invention, it is proposed to introduce in the encoding syntax
used by the video standards and recommendations an additional information consisting
of a new syntactic element supporting their lack of flexibility and opening new
possibilities to encode more efficiently and independently the temporal prediction of

3

various channels. This additional syntactic element, called for example "channel temporal prediction", takes the following symbolic values :

    Motion_compensation

    Temporal_copy

    Temporal_interpolation

    Motion_compensated_temporal_interpolation,

and the meaning of these values is :

    a) motion_compensation : the temporal prediction is formed by directly applying the motion field sent by the encoder on one or more reference pictures (this default mode is implicitly the INTER coding mode of most of the current coding systems) ;

    b) temporal_copy : the temporal prediction is a copy of a reference image ;

    c) temporal_interpolation : the temporal prediction is formed by the temporal interpolation of the motion fields ;

    d) motion_compensated_temporal_interpolation : the temporal prediction is formed by the temporal interpolation of the current motion field and further refined by the motion field sent by the encoder.

    The words "temporal interpolation" must be understood in a broad sense, i.e. as meaning any operation of the type defined by an expression such as $Vnew = a.V1 + b.V2 + K$, where V1 and V2 designate previously decoded motion fields, a and b designate coefficients respectively assigned to said motion fields, K designates an offset and Vnew is the new motion field thus obtained. It can therefore be seen that, in fact, the particular case of the temporal copy is included in the more general case of the temporal interpolation, for $b = 0$ and $K = 0$ (or $a = 0$ and $K = 0$).

    According to the invention, the additional syntactic element thus proposed has to be placed at the following levels in the coded bitstream that has to be stored (or to be transmitted to the decoding side):

1)   either at the slice level ;

2)   or at the macroblock level ;

this additional syntactic element being in each case either specific for each present channel or, possibly, shared by all the channels.

    This invention may be used in some identified situations where the way of constructing the temporal prediction can be switched on a slice or macroblock basis, and also on a channel basis. A first example may be for instance a sequence with a shape channel : it is possible that the shape information does not change much, whereas the

luminance and chrominance channels carry varying information (it is for instance the case with a video depicting a rotating planet : the shape is always a disc, but the texture of it depends on the planet rotation). In this situation, the shape channel can be recovered by temporal copy, and the luminance and chrominance channels by motion compensated temporal interpolation. A second example may be the case of a change at the macroblock level. In a video sequence showing a seascape with the sky in the upper part of the picture, unlike the sea, the sky remains the same from one image to the other. Its macroblocks can therefore be encoded by temporal copy, whereas the macroblocks of the sea have to be encoded by temporal interpolation.

5

CLAIMS :

1.   An encoding method applied to an input video sequence corresponding to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of said scenes, a coded bitstream the content of which is described in terms of separate channels and constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of said content, said syntax comprising an additional syntactic information provided for describing independently the type of temporal prediction of the various channels, said predictions being chosen within a list comprising the following situations :

             - the temporal prediction is formed by directly applying the motion field sent by the encoder on one or more reference pictures ;

             - the temporal prediction is a copy of a reference image ;

             - the temporal prediction is formed by the temporal interpolation of the motion field ;

             - the temporal prediction is formed by the temporal interpolation of the current motion field and further refined by the motion field sent by the encoder ; said method being further characterized in that said additional syntactic information is a syntactic element placed at the slice level in said generated coded bitstream and its meaning is specific for each present channel.

2.   An encoding method applied to an input video sequence corresponding to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of said scenes, a coded bitstream the content of which is described in terms of separate channels and constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of said content, said syntax comprising an additional syntactic information provided for describing independently the type of temporal prediction of the various channels, said predictions being chosen within a list comprising the following situations :

             - the temporal prediction is formed by directly applying the motion field sent by the encoder on one or more reference pictures ;

             - the temporal prediction is a copy of a reference image ;

             - the temporal prediction is formed by the temporal interpolation of the motion field ;

- the temporal prediction is formed by the temporal interpolation of the current motion field and further refined by the motion field sent by the encoder ; said method being further characterized in that said additional syntactic information is a syntactic element placed at macroblock level in said generated coded bitstream and its meaning is specific for each present channel.

3.   An encoding method according to anyone of claims 1 and 2, characterized in that said meaning is shared by all existing channels.

4.   An encoding device processing an input video sequence that corresponds to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of said scenes, a coded bitstream the content of which is described in terms of separate channels and constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of said content, said encoding device being provided for carrying out the encoding method according to anyone of claims 1 and 2.

5.   A transmittable video signal consisting of a coded bitstream generated by an encoding device processing an input video sequence that corresponds to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of said scenes, a coded bitstream the content of which is decribed in terms of separate channels and constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of said content, said transmittable video signal including an additional syntactic information provided for describing independently the type of temporal prediction of the various channels, said predictions being chosen within a list comprising the following situations :

         - the temporal prediction is formed by directly applying the motion field sent by the encoder on one or more reference pictures ;

         - the temporal prediction is a copy of a reference image ;

         - the temporal prediction is formed by the temporal interpolation of the motion field ;

         - the temporal prediction is formed by the temporal interpolation of the current motion field and further refined by the motion field sent by the encoder ; and said additional syntactic information being a syntactic element placed at the slice level or at the macroblock level in said generated coded bitstream and its meaning is specific for each present channel.

6.   A method for decoding a transmittable video signal consisting of a coded bitstream generated by an encoding device processing an input video sequence that corresponds to successive scenes subdivided into successive video object planes (VOPs) and generating, for coding all the video objects of said scenes, a coded bitstream the content of which is described in terms of separate channels and constituted of encoded video data in which each data item is described by means of a bitstream syntax allowing to recognize and decode all the elements of said content, said transmittable video signal including an additional syntactic information provided for describing independently the type of temporal prediction of the various channels, said predictions being chosen within a list comprising the following situations :

        - the temporal prediction is formed by directly applying the motion field sent by the encoder on one or more reference pictures ;

        - the temporal prediction is a copy of a reference image ;

        - the temporal prediction is formed by the temporal interpolation of the motion field ;

        - the temporal prediction is formed by the temporal interpolation of the current motion field and further refined by the motion field sent by the encoder ; and said additional syntactic information being a syntactic element placed at the slice level or at the macroblock level in said generated coded bitstream and its meaning is specific for each present channel.

7.   A decoding device for carrying out a decoding method according to claim 6.

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7   H04N7/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
IPC 7   H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	"MUNICH MEETING OF MPEG-4 WORKING GROUP. REPORT ISO/IEC JTC1/SC29/WG11 MPEG4/N1172" INTERNATIONAL ORGANIZATION FOR STANDARDIZATION - ORGANISATION INTERNATIONALE DE NORMALISATION, XX, XX, 1996, pages 3-49, XP002047798 page 5; figure 3.1.2 page 11, paragraph 3.3.2 - page 15, last line ——— -/--	1-7

[X] Further documents are listed in the continuation of box C.	[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 July 2004	02/08/2004

Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Schoeyer, M

Form PCT/ISA/210 (second sheet) (January 2004)